

## Random forests algorithm in podiform chromite prospectivity mapping in Dolatabad area, SE Iran

M.P. Sadr\* and M. Nazeri

*Applied Research and Independence Center of Khatamolanbia Company, Tehran, Iran*

Received 4 March 2018; received in revised form 25 March 2018; accepted 30 March 2018

\*Corresponding author: muhamadmilresearch@gmail.com (M.P. Sadr).

### Abstract

The Dolatabad area located in SE Iran is a well-endowed terrain owning several chromite mineralized zones. These chromite ore bodies are all hosted in a colored *mélange* complex zone comprising harzburgite, dunite, and pyroxenite. These deposits are irregular in shape, and are distributed as small lenses along colored *mélange* zones. The area has a great potential for discovering further chromite resources. Therefore, the current work endeavors to delineate the favorable zones of podiform chromite mineralization to focus on the detailed exploration surveys. In order to achieve this goal, the machine learning random forests algorithm was adapted to integrate the footprints of mineralization in various exploration datasets. The genetic characteristics of podiform chromite deposits were used to define the exploration criteria. These defined criteria were then translated to a set of exploration evidence layers. The competent exploration evidence layers, i.e. those with remarkable positive spatial associations with mineralization, were then recognized using distance distribution analysis. Respecting the location of known chromite mineralizations and competent exploration evidence layers, a predictive random forests model was trained and then applied to predict the favorable zones of chromite prospectivity. The delineated targets were found to occupy 19% of the studied area, in which all the known chromite mineralizations were delimited. Consequently, it is worthy to follow up the detailed exploration surveys within the delineated zones.

**Keywords:** *Podiform Chromite Deposits, Random Forests, Mineral Prospectivity Mapping.*

### 1. Introduction

Due to the increase in metal consumption, depleted mineral reserves should be replaced by newly discovered mineral resources. Therefore, mineral exploration surveys should focus on the target zones of mineral prospectivity, in which mineral endowments are highly probable to be discovered. In regional-scale exploration surveys, the exploration targets are delineated by fusion of geo-exploration signatures of mineral deposits of the type sought or the so-called mineral prospectivity modeling (MPM) [1-17]. MPM is a multi-step process comprising (a) selection of the exploration criteria of the deposit-type sought, (b) translation of the exploration targeting criteria to a set of competent 2D exploration evidence layers, (c) weighting and integration of the individual exploration evidence layers, and finally, (d)

prioritization of the target zones for detailed exploration surveys [2].

Generation of the competent 2D exploration evidence layers in regional-scale exploration surveys is a challenge. This is because in these terrains, not only the quality of the available datasets might be questionable but also the link between these datasets and the deposit-type sought is not clear [2, 18]. The competent exploration evidence layers are those with remarkable positive spatial associations with deposits of the type sought [2, 12, 18]. Therefore, the competent exploration evidence layers should be recognized from the incompetent ones. For this, the spatial association between the exploration evidence layers and the known mineral deposits of the targeted type should be

measured [2, 18, 19]. Different methods have been applied to assess the spatial distribution of the exploration evidence layers and the known mineral deposits including studentized t-statistic value [1], success-rate curves [20], receiver operating characteristics curves [21], and distance distribution analysis [22].

Weighting and synthesis of the exploration evidence layers is another challenging step of MPM. This is due to the non-linear nature of ore-forming processes [23], which should be modeled with proper tools [21, 24]. There are knowledge- and data-driven methods for weighting and synthesis of the exploration evidence layers [1-2]. The former and the latter are appropriate to be applied in the data-poor and data-rich zones, respectively [2]. Theoretically, multi-variate non-linear data-driven methods are more appropriate for modeling the non-linear nature of ore-forming processes [25]. Recently, machine learning algorithms have gained a considerable reputation in the data-driven modeling of mineral prospectivity [5, 7, 15, 25]. These methods are multi-variate non-linear algorithms, which can properly model the complexity of geological processes [25]. Random forests algorithm [26-28], as a machine learning method, has been successfully adapted in data-driven MPM [25, 29-31]. The superiority of random forests over other machine learning algorithms has been demonstrated in data-driven MPM [25, 30]. Therefore, it is worthy to apply the random forests algorithm in data-rich terrains to delineate the plausible zones of mineral prospectivity.

The Dolatabad area is a well-endowed terrain in SE Iran owning more than 20 podiform chromite deposits [32]. Several geologists have reached a consensus about the great potential of the Dolatabad area for prospecting a further chromite mineralization [33]. The main objective of this work, therefore, is to delimit the target zones of podiform chromite deposits in this wealthy area. For this, a set of exploration datasets comprising the output of remote-sensing studies for detecting the hydrothermal alteration zones, and the geological and geochemical data was used as an exploration dataset, from which a set of exploration evidence layers were generated. The competent exploration evidence layers were then recognized by distance distribution analysis [22] and used as predictor variables in machine learning random forests algorithm.

## 2. Dolatabad area

### 2.1. Geological setting

The Dolatabad area is situated in the southern part of the Sanandaj–Sirjan metamorphic zone, and covers an area of 1500 km<sup>2</sup>. The area is also a part of the Dolatabad-Esfandagheh ophiolite complex in southern Iran (Figure 1). This ophiolite complex is Cenozoic-Mesozoic in age, and serves as the host rock of several mineralization types for different commodities such as Cr, Mg, Mn, and Cu [33].

This area is covered by the outcrops of Paleozoic mafic and ultramafic sequences comprising pyroxenite, peridotite, gabbro, dunite, and harzburgites, followed by the mafic-ultramafic metamorphosed rocks of Paleozoic and Mesozoic age (Figure 2). These metamorphosed rocks are glaucophane schist, serpentine schist, amphibolite, and greenschist, all of which are associated with chromite mineralization and ultramafic sequences [32-33]. The Dolatabad-Esfandagheh ophiolitic colored mélange complex covers a considerable part of the area, which is mainly covered by massive oolitic and pelagic limestones, basalts, andesite basalts, and dacites. These basalts envisage the occurrence of pillow lava flows in the ophiolitic sequences, and are partly affected by the argillic and phyllic alterations [32-33]. There are also outcrops of intermediate to basic plutonic bodies, which are Late Triassic gabbros and Cenozoic granodiorites. Jurassic limestones and conglomerates and also alternations of Cretaceous shale, sandstone, and limestones are dominated in the NW part of the area. The lower slopes of the area are mainly covered by the Quaternary alluvial deposits (Figure 2). This figure shows the generalized geological map of the Dolatabad area.

### 2.2. Podiform chromite mineralization

The podiform chromite deposits are among the crucial resources of chromite, which is the only economic mineral for chromium [35]. These deposits are formed as small magmatic lenses within the ultramafic section of colored mélange and ophiolite sequences in oceanic crusts [36-39]. They are usually formed in supra-subduction tectonic settings [39], meaning that they were a part of the oceanic crust that arose to the surface during the subduction of oceanic crust below the continental crust, most probably due to their specific weight [39]. Podiform chromite mineralization is hosted by dunite, serpentine or peridotite bodies [36]. The peridotite host rocks include harzburgite and lherzolite. Chromitite in

the peridotite bodies is almost associated with dunite or serpentine bodies that are near gabbro [37] or pyroxenite [36]. Some podiform chromite deposits are accompanied by a halo of serpentinized alteration [38-40]. The podiform chromite deposits are often affected by faulting and tectonic processes [39]. Most of these deposits have been displaced by the interaction

between faults, and consequently, their size has been significantly reduced [41]. There are 29 podiform chromite occurrences in the Dolatabad area, most of which are accompanied by thin serpentinized halos and hosted by the donite and peridotite zones [33]. These deposits were used to generate a set of training data in order to develop the predictive model.

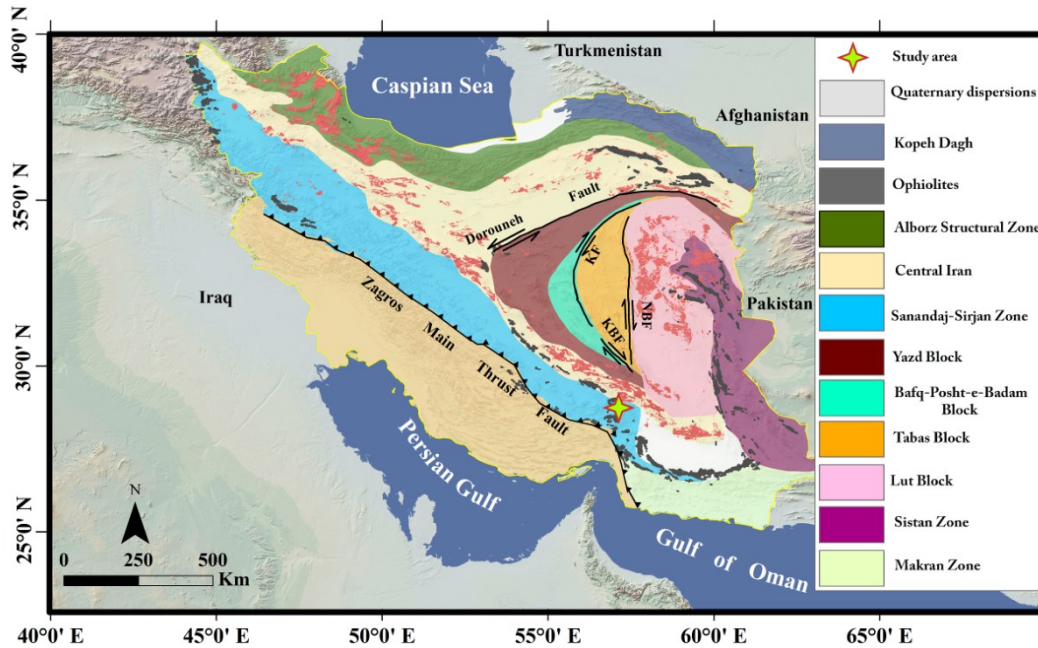


Figure 1. Location of Dolatabad area in Sanandaj-Sirjan zone [34].

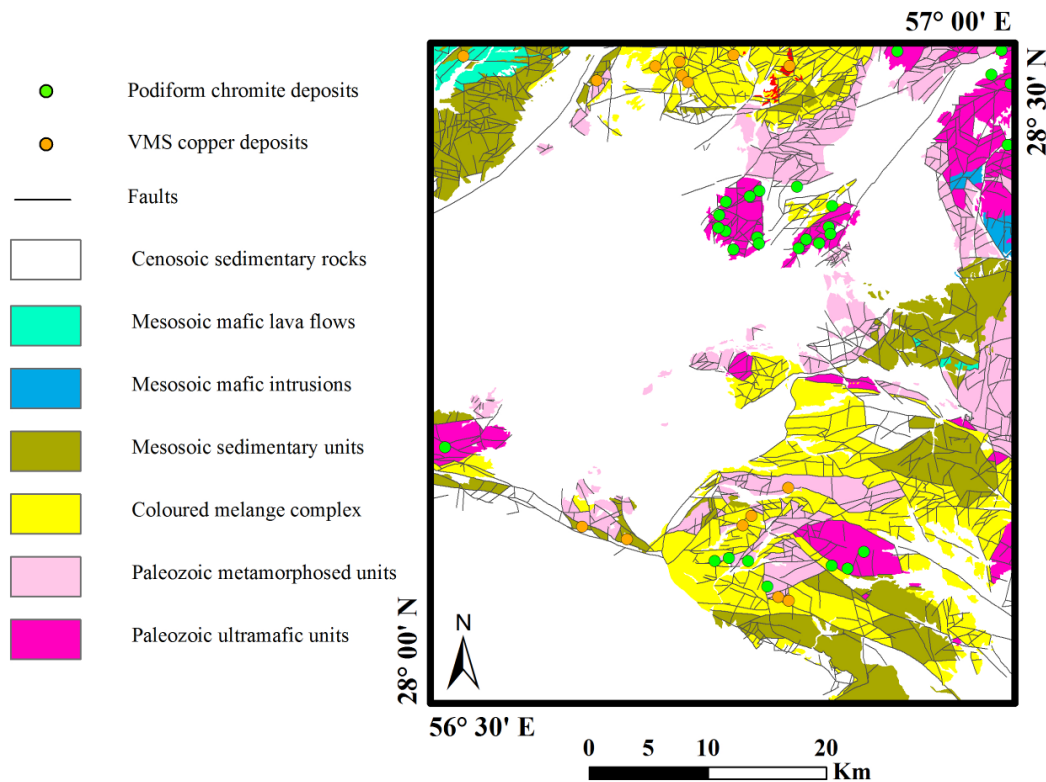


Figure 2. Simplified geological map of Dolatabad area [33].

### 2.3. Dataset

The data used in this work included (a) a 1:100000 geological map of the Dolatabad area, from which the lithological units and faults had been extracted; (b) the results of processing of Aster remotely-sensed data for extraction of the serpentinized altered zones; and (c) the results of chemical analyses of the stream sediment geochemical samples for Cr, Co, Ni, Au, Zn, Pb, and Cu, collected by the geological survey of Iran (GSI). These stream sediment samples were analyzed by inductively coupled plasma optical emission spectrometry. However, gold was separately assessed by the fire assay method. The detection limits for the analyzed elements were 0.2 ppm for Co, Pb, Zn, and Cu; 2 ppm for Cr and Ni; 0.5 ppm for As; and 1 ppb for Au. The analytical precision of these elements was assessed by the Thompson-Howarth method (1976) [42]. According to this method, the analytical precision of the analyzed elements was better than 10% at the 95% of confidence level.

## 3. Methods and results

### 3.1. Generation of exploration evidence layers

According to the genetic model of the podiform chromite deposits [39], the following exploration criteria were selected for MPM: (a) dunite, serpentine, and peridotite bodies are the host rock of podiform chromite deposits [36-41]; (b) some of the podiform chromite deposits could experience different phases of serpentinized alteration [39]; (c) in regional-scale studies, multi-variate analyses could be applied to the stream sediment geochemical data for the recognition of mineralization-related geochemical signatures [8, 9, 11]; and (d) some faults are constructive in the development and exposure of the podiform chromite deposits [39].

The selected exploration criteria should be transferred to the 2D mappable features or evidence layers to be weighted and integrated. In this work, the exploration evidence layers were continuously generated to modulate the systematic uncertainty, which resulted from the fallacious discretization of evidence layers [12-14]. A pixel size of 200 m  $\times$  200 m was used to generate all the exploration evidence layers and the rest of processes in MPM. This pixel size was selected according to an objective procedure outlined by Carranza (2009) [43]. In the following sub-sections, the generation of host rock, alteration, fault density, and mineralization-related geochemical evidence layer is explicitly explained.

#### 3.1.1. Evidence layer of host rock

The distances to the polygons of dunite, serpentine, and peridotite bodies digitized from the published 1:100,000 geological map of the studied area by the geological survey of Iran [32] was calculated and mapped using a GIS system as the evidence for host rock (Figure 3).

#### 3.1.2. Evidence layer of alteration

The distances to the polygons of serpentinized alteration detected by the processing of ASTER remotely-sensed data were calculated and mapped using a GIS system as the evidence for mineralization-related alteration (Figure 4).

#### 3.1.3. Evidence layer of faulting

Faults could be constructive or destructive in the development and exposure of the podiform chromite deposits [38]. Therefore, it is worthy to generate an exploration evidence layer associated with faulting, and assess its spatial association with the known podiform chromite deposits. To this end, the fault density layer was generated (Figure 5) using the calculation of the total length of faults per pixel in this area [12].

#### 3.1.4. Evidence layer of geochemistry

Aiming to identify a mineralization-related multi-element geochemical footprint of mineralization, the robust factor analysis of compositional data [44] was used [e.g. 45, 46]. Robust factor analysis is a modified version of the ordinary factor analysis through which two problems of this analysis have been modulated. These problems are (a) the adverse effect of extreme values in factor estimation, and (b) the closure problem of geochemical data, which renders bias through estimation of factor loadings and scores [44]. The robust factor analysis of compositional data modulates the former problem using robust estimators by which extreme values no longer affect the factor estimation [44]. Moreover, the latter problem is addressed using log-ratio transformations such as the isometric log-ratio [ilr: 47] and centered log-ratio [clr: 48] transformations. The details of this method could be found in the related publications [44]. In the robust factor analysis of compositional data, the relationship between the estimated factors and elements is described by factor loadings, while the relationship between the estimated factors and samples is described by factor scores [44].

Initially, the raw data for elements were subjected to ilr transformation [47] in order to address the data closure problem [48]. Then the robust factor

analysis of compositional data [44] was applied to the ilr-transformed data to derive the loading and score matrices. Since the ilr transformation does not yield a one-to-one representation of the transformed variables [47], the derived loading and score matrices were back-transformed to the clr space [48], in which the interpretation of results is possible using the loading matrix [44]. Three factors were selected to be extracted from the robust factor analysis. Aiming to interpret the retained factors, i.e. the extracted factors, a value between 0.4 and 0.9 should be used, above which the contribution of elements in a factor can be assumed as significant [45, 46, 49]. In this work, the absolute value of 0.5 was used for interpretation of the retained factors because this value not only allows for the contribution of moderately high and moderately low values in the interpretation of factors but also has succeeded in the discrimination between the mineralized and

non-mineralized element associations in this work [15, 45, 46, 49]. The first and second factors describe 45.56% and 26.76% of the total variability of the dataset, respectively (Table 1). The first component has positive loadings for Cr, Ni, Co, Zn, and Cu, and represents a Cr-Ni-Co elemental association (Table 1). This factor probably describes the ultramafic and ophiolitic rock sequences [8, 9, 11]. The foregoing factor could be considered as a multi-variate geochemical signature of the deposit-type sought [8, 9, 11]. Due to a straightforward implementation, the inverse distance weighting (IDW) method was applied for modeling the spatial distribution of the component scores [12]. Figure 6 depicts a map of the spatial distribution of the first component scores. This map was used as an evidence layer of the podiform chromium deposits in this work.

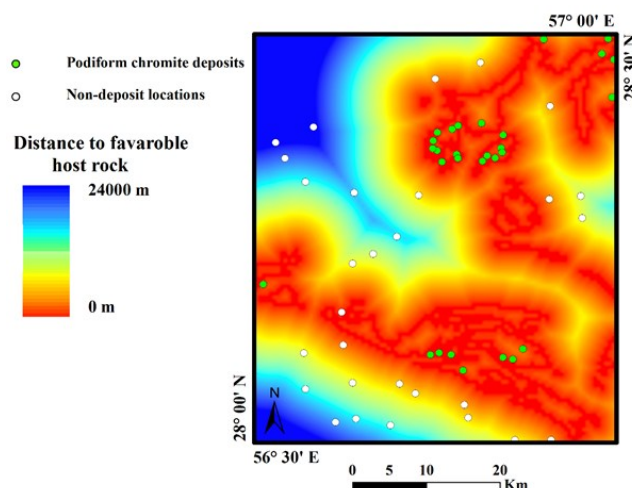


Figure 3. Evidence of host rock for podiform chromite mineralization in the studied area.

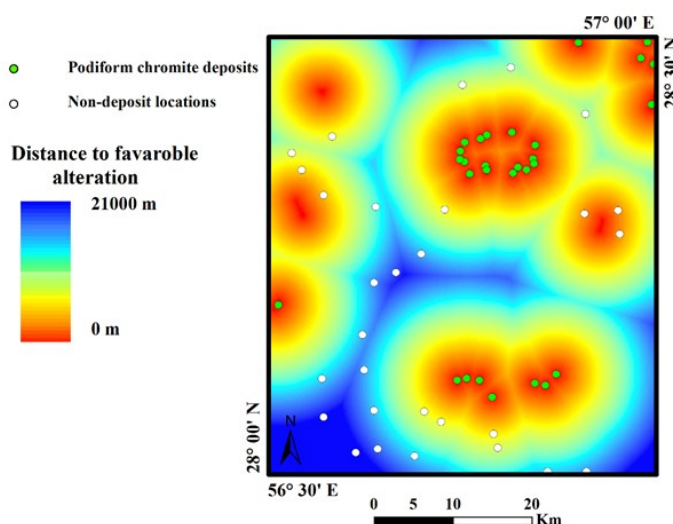


Figure 4. Evidence of serpentinized alteration for podiform chromite mineralization in the studied area.



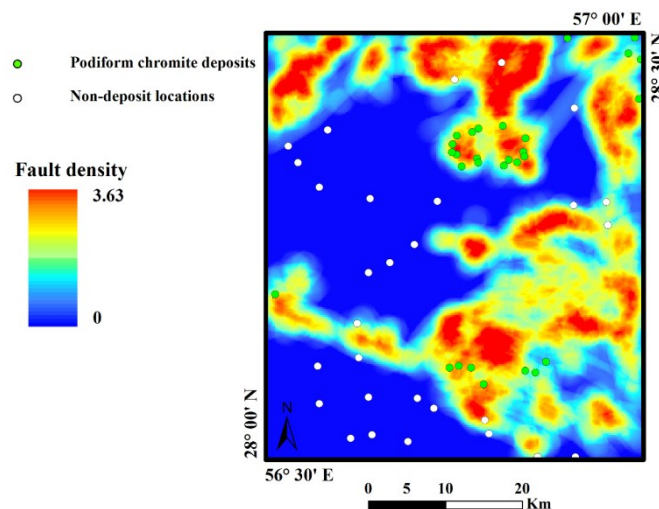


Figure 5. Exploration evidence layer of fault density.

Table 1. Loading matrix of three retained factors. Significant loadings (the bolded values) were selected based on the absolute threshold value of 0.5 [47].

Element	Robust factor 1	Robust factor 2	Robust factor 3
Cu	0.12	-0.21	<b>0.55</b>
Au	-0.43	<b>0.51</b>	-0.23
As	-0.21	0.46	-0.45
Pb	-0.12	<b>-0.61</b>	-0.32
Zn	0.15	-0.18	<b>0.61</b>
Cr	<b>0.54</b>	0.12	0.20
Co	<b>0.63</b>	0.09	0.18
Ni	<b>0.51</b>	0.11	0.23
Variability	45.56%	26.76%	11.25%
Total variability	45.56%	72.32%	83.57%

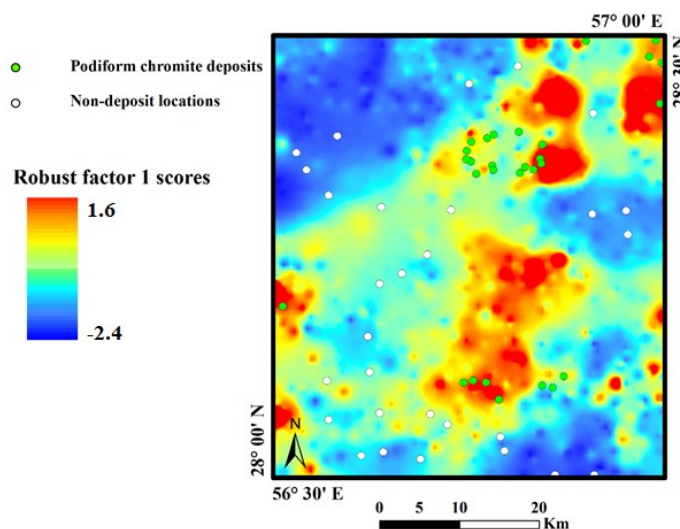


Figure 6. Spatial distribution of the first robust factor representing a Cr-Ni-Co elemental association.

### 3.2. Recognition of competent exploration evidence layers

Distance distribution analysis [22] is used to measure the spatial association between the mineralization and geological features [2, 17, 19]. Two curves are simultaneously constructed in this analysis comprising (a) distance to every location

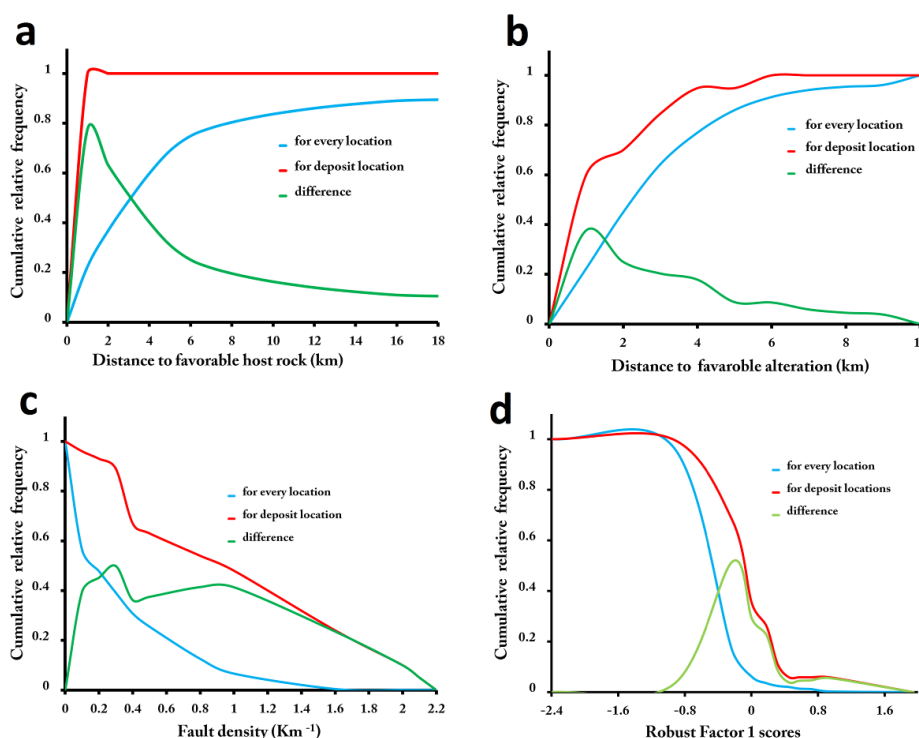
and (b) distance to deposit locations. The former and the latter curves are representatives of the random and ore-forming processes, respectively [19]. The difference between the latter and the former curves shows the degree of spatial association between the mineral deposit locations and the geological features. This difference is

shown in a contrast curve, in which positive and negative values show positive and negative spatial associations between deposit location and geological features, respectively. The competent exploration evidence layers, for which the contrast curve lies in a positive range, could be recognized by the distance distribution analysis [19].

The distance distribution curves [22] were plotted for recognition of the competent exploration evidence layers. The closer to the host rock and alteration is, the more favorable for mineralization would be. Therefore, cumulative increasing distance distribution curves [19] were generated for these two evidence layers (Figs. 7a,b). However, the higher the geochemical values and

fault density values are, the more favorable for mineralization would be. Therefore, the cumulative decreasing distance distribution curves [19] were generated for these two evidence layers (Figs. 7c,d).

All the contrast curves in Figure 7 show remarkable positive values, and therefore, the spatial association between the generated evidence layers and the known podiform chromite deposits is positive. Consequently, it should be concluded that all the generated exploration evidence layers are competent [19] and can be used in targeting the prospective zones of podiform chromite deposits in the Dolatabad area.



**Figure 7. Distance distribution analysis for three generated evidence layers: (a) host rock, (b) alteration, (c) fault density, and (d) geochemical evidence layers.**

### 3.3. Predictive modeling by random forests algorithm

#### 3.3.1. Random forests algorithm

Random forests are ensembles of multiple decision trees [26] that could be applied in the regression and classification problems [26-28]. Each decision tree uses a random subset of training samples that are taken with replacement from the original data [26-28]. About two-thirds of the samples are used for training the model, while the rest of them that are called the out-of-bag (OOB) samples are used for validation of the results using the OOB error [26-28].

The random forests algorithm models a target

variable (here, the probability of occurrence of mineralization) based upon a set of predictor variables (here, exploration evidence layers). Random forests algorithm starts with purification of child nodes through splitting the target variable based on predictor variables from the parent node. The splitting successively iterates until a pre-defined stop criterion is reached. Through this process, every decision tree reaches its simple regression or classification model. Random forests algorithm then averages the results of various decision trees to gain the final model. The details of the random forests algorithm could be found in the related publications [26-28].

Generally, the implementation of random forests algorithm in data-driven MPM includes the following steps [25, 29-31]:

1. Generation of training data, which include the values for the competent exploration evidence layers in the location of deposit and non-deposit locations as the positive and negative sites, respectively.
2. Training a set of random forests based on the training data. In order to train a set of random forests, a number of parameters should be set. These parameters are (a) the number of decision trees to be used, and (b) the number of predictor variables (evidence layers) to enter each decision tree.
3. The trained random forests are then used for predictive mapping. The competent exploration evidence layers are used as the predictors or independent variables in the trained random forests, based on which, the probability of occurrence of mineralization as the target variable or dependent variable is estimated.
4. The results of predictive modeling are imported to GIS or any other 2D visualization software for mapping the exploration targets. The aforementioned steps are explained in the following sub-sections.

### 3.3.2. Generation of training data

Similar to all the predictive machine learning algorithms, random forests algorithm requires the training data to derive the predictive model [26-28]. The training data comprises the values of evidence layers for the location of mineral deposits and also non-deposit locations as the positive and negative sites, respectively [29-31]. The former can be achieved by the spatial distribution of mineral deposits, while the latter should be derived by a series of analyses, which are explained as follow [19, 21]. Firstly, the number of non-deposit locations should be equal to the deposit locations. There are 29 podiform chromite deposits in the studied area, and therefore, the number of non-deposit locations should be 29. The non-deposit locations should be far from the deposit locations so that their geological characteristics would be different from the deposit locations. Therefore, point pattern analysis [22] was applied to recognize the least possible distance from the deposit locations that suit non-deposit locations. This distance was recognized to be 5 km from the deposit locations. Moreover, despite the deposit locations that usually have a clustered distribution, the

non-deposit locations should be randomly distributed [2]. The above-mentioned procedure was adapted for the generation of 29 randomly-distributed non-deposit locations in the Dolatabad area (Figures 3-6). The values for evidence layers or predictor variables were extracted for the location of podiform chromite deposits and also non-deposit locations. These values were used as the training data for running the predictive model.

### 3.3.3. Predictive modeling

The training data was subjected to the "randomForests" package [50] in R freeware [51] for training a predictive regression model. As mentioned earlier, a couple of parameters are required to be rectified before running the regression model, namely the number of regression trees to be used in the model and the number of predictor variables to be incorporated in each regression tree of the "randomForests" package [50]. In this work, the number of regression trees was selected to be 20,000 since picking a large number of regression trees could enhance the efficiency of the training procedure [29-31]. Moreover, the number of predictor variables to enter each node was selected using the "tuneR" function [29-31]. The trained model was then used to predict the prospectivity scores of different pixels in the Dolatabad area. Figure 8 depicts a map of the podiform chromite prospectivity in the Dolatabad area.

The importance of evidence layers contributed to the generated model was then assessed using the GINI impurity index [28]. The GINI impurity index is a measure of how each variable contributes to the homogeneity of the nodes, and involves the generated random forests. Each time a variable is used to split a node, the GINI impurity index for the child nodes is calculated and compared to that of the original node. The GINI impurity index is a measure of homogeneity from 0 (homogeneous) to 1 (heterogeneous). The changes in GINI are summed for each variable and normalized at the end of the calculation. The variables that result in nodes with higher purity have a higher GINI index. The higher the GINI index for a variable is, the more important the variable in the predictive modeling would be [28]. This index demonstrates that the most important evidence layer in podiform chromite prospectivity mapping is host rock, while the least important evidence layer is the geochemical signature (Figure 9).



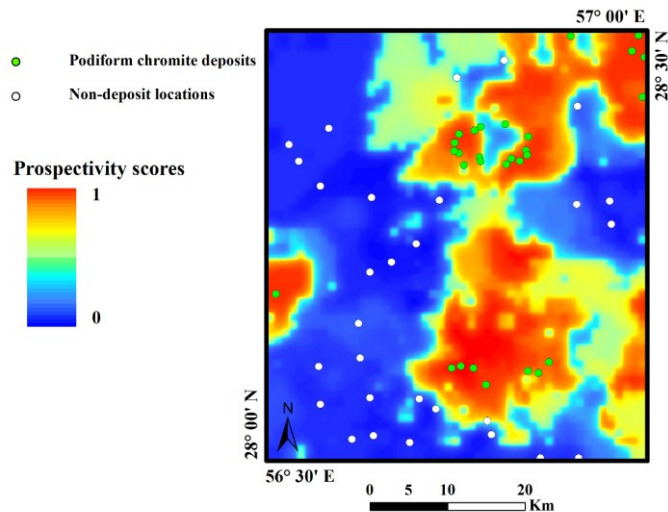


Figure 8. Predictive model of podiform chromite prospectivity.

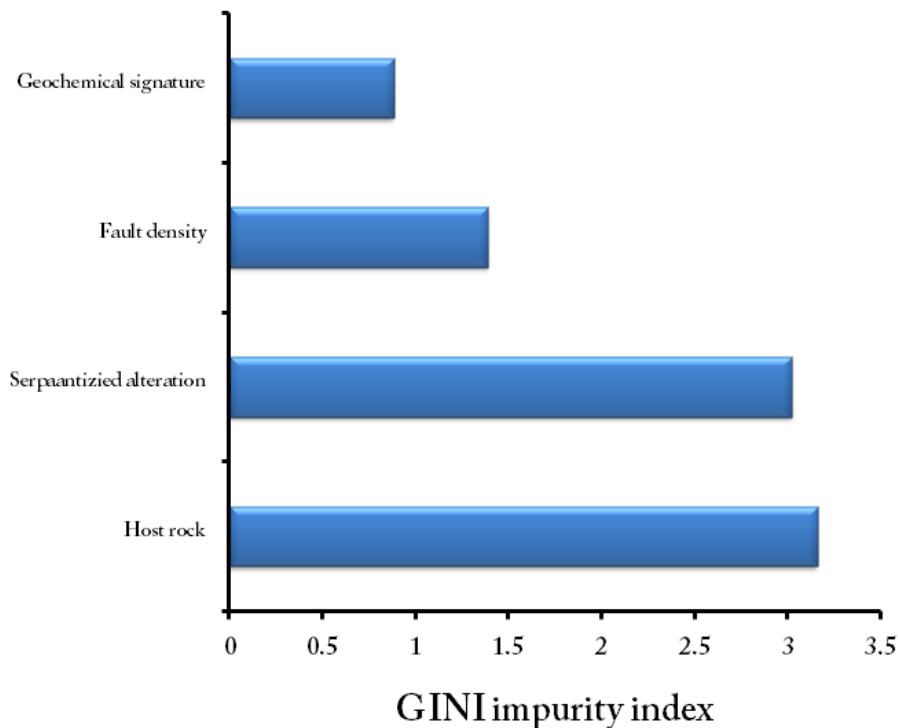


Figure 9. Importance of exploration evidence layers contributed to the predictive model of podiform chromite prospectivity.

The values for prospectivity mapping by random forests algorithm lie in the [0, 1] range. Therefore, the prospectivity values  $> 0.5$  and the values  $< 0.5$  could be considered as the prospective and non-prospective values, respectively [29-31]. According to the mentioned premise, a table of

correctness [29-31] was generated to assess the accuracy of the derived prospectivity model. According to this table, the error for classification of the deposit and non-deposit locations is 0 (Table 2).

Table 2. Table of correctness for assessing the accuracy of the generated predictive model.

Error for classification of deposit locations	0 out of 29 (0%)
Error for classification of non-deposit locations	0 out of 29 (0%)
OOB error	0%
Model error	0%

The success- [20] and prediction-rate curves [52] could be employed to verify the predictive models. The former measures the correlation of the generated model with known deposit locations, while the latter estimates the chance of discovering further mineral endowments [29-31]. For the generation of success-rate curves, the prospectivity scores should undergo a series of discretization processes based on the cut-off values at five percentile intervals. Based upon each cut-off value, the two parameters (a) the portion of the area enclosed by the pattern representing the delimited prospectivity zone ( $P_a$ ) and (b) the portion of mineral deposits delimited in that prospectivity zone ( $P_d$ ) are calculated. The values for  $P_a$  and  $P_d$  for different threshold values are estimated and then plotted in the vertical and horizontal axes, respectively [20].

For construction of the prediction-rate curves, the leave-one-out approach [52] was applied. In this approach, one deposit is excluded from a set of  $m$  deposits, and the remaining  $m-1$  deposits are used for generating a prospectivity model. Then the generated prospectivity model is cross-validated with the excluded deposit. This procedure iterates  $m$  times, each time with excluding a different deposit for cross-validating. The values for mineral prospectivity models at each excluded deposit are used as thresholds for classification of

mineral prospectivity scores. The proportion of delimited prospectivity areas by the  $m$  threshold values in the  $m$  generated prospectivity models are sorted increasingly. Then the cumulative increasing proportion of areas ( $P_a$ ) and the cumulative increasing proportion of the deposits ( $P_d$ ) are derived. Similar to the success-rate curves, the prediction-rate curves are constructed using the  $P_a$  values on the horizontal axis versus the  $P_d$  values on the vertical axis [52]. For a reliable predictive model, the success- and prediction-rate curves must lie above the diagonal line, which connects the first and last points of the curves [20, 53]. This is because the diagonal line represents a completely random process, and could not be accounted as a reliable prediction [20, 53]. Besides, as it has been discussed in several publications [15, 21, 25, 29-31], the success-rate curve must always lie above the prediction-rate curve. Figure 10 shows the generated success- and prediction-rate curves. The model is generally good because the success- and prediction-curves lie above the diagonal line (Figure 10). Moreover, the model is correctly generated as its success rate lies above its prediction rate [29-31]. Therefore, the generated model could be used to define the target areas for detailed exploration surveys.

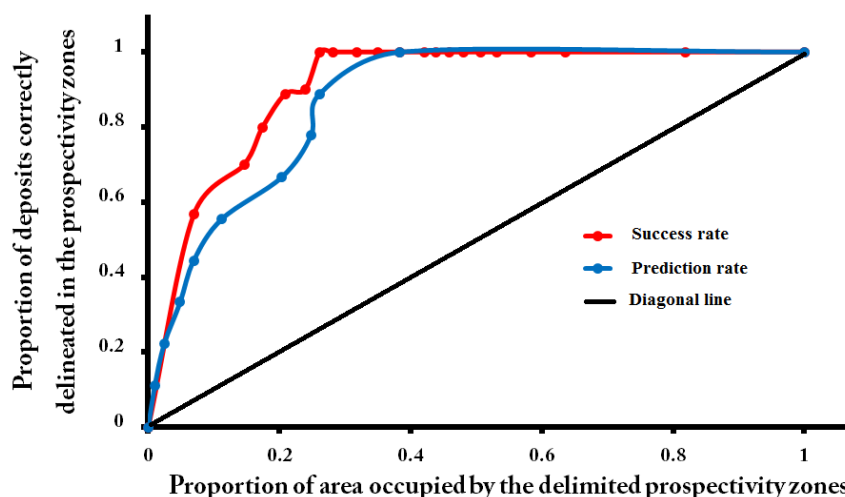


Figure 10. Success- and prediction-rate curves for the generated prospectivity models.

### 3.3.4. Exploration targets

Discretization of continuous prospectivity models can be implemented either subjectively or objectively [25, 29-31]. In this work, the subjective value of 0.85 was used for discretization of the continuous prospectivity model of Figure 8. It means that the exploration

targets are pixels whose prospective scores have the values  $> 0.85$ . Figure 11 demonstrates the derived exploration targets. These targets have occupied 19% of the studied area, while covering all the known chromite deposits. It demonstrates that the generated targets are reliable to be followed up by further exploration surveys.

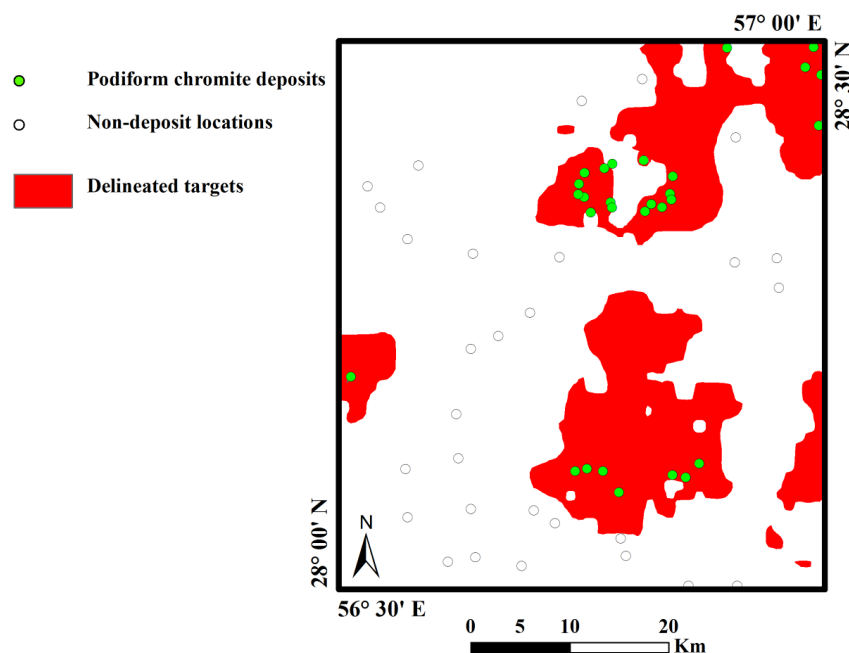


Figure 11. Delimited exploration targets for podiform chromite mineralization.

#### 4. Discussion

Mineral prospectivity mapping (MPM) is plagued by different types of systematic uncertainties [2]. The major sources of systematic uncertainties in the regional-scale MPM [12-14, 18, 24] are as follow: (a) fallacious generation (b) improper weighting and integration of the exploration evidence layers. The former case is due to the failures and inherent errors in the exploration datasets, which yield an unclear link between the exploration datasets and the deposit-type sought [18]. In order to address the foregoing uncertainty, the competent exploration evidence layers should be recognized from the incompetent ones by assessing their spatial association with known deposits of the type sought [12-14, 18]. In this work, distance distribution analysis [22] was applied to assess the spatial association between the geological and geochemical evidence layers and mineralization. This procedure not only lowers the uncertainties but also enhances the predicting ability of MPM [12-14]. Multi-variate non-linear algorithms can be utilized to modulate the second source of systematic uncertainty in MPM [25]. In this work, random forests algorithm [29-31] was employed for weighting and integration of the competent exploration evidence layers. The merits of this method are as follow: (a) its non-linear combination of predictor variables (b) its ability to use continuous exploration evidence layers (c) its superiority over lots of machine learning algorithms [25]. The applied procedure in this work comprising the selection of

competent exploration evidence layers using spatial analyses and application of random forests algorithm for data-driven MPM worked well as it could be understood by the accuracy of the prospectivity model (Table 2) and its success and prediction rates (Figure 10). Therefore, this procedure could be adapted in different case studies for delineation of reliable exploration targets.

The recognized exploration targets in this work are mostly correlated with ultramafic and mafic rock sequences, which are the host of chromite mineralization in SE Iran [33]. It was not only geologically expected but also the algorithm did well in the recognition of the most important exploration evidence layers (Figure 9). It is worthy to note that few parts of ultramafic rock sequences of the Dolatabad area are mineralized [33]. Therefore, the target zones of the podiform chromite deposits were recognized by the integration of the geochemical, faulting, host rock, and alteration evidence layers. Exploration licenses and mining tenements for chromite deposits can be followed according to the recognized exploration targets (Figure 11). Large-scale detailed exploration surveys for the podiform chromite deposits can then be implemented in the licensed tenements within the delimited target zones. However, large-scale exploration surveys for the podiform chromite deposits have been a great challenge [54]. This is mainly due to the unpredictable nature of these deposits [36-38]. Field geological surveys are the

best approaches in large-scale exploration of the podiform chromite deposits [55]. The results of field geological surveys are worthy to be facilitated with gravimetric geophysical surveys [55]. Therefore, it is recommended to follow-up mixed geological and gravimetric geophysical surveys in the generated exploration targets.

## 5. Conclusions

In this work, we demonstrated the application of random forests algorithm for prediction of the target zones of podiform chromite. The results of processing remote sensing, geological, and geochemical data were used as inputs for mineral prospectivity mapping. A set of continuous exploration evidence layers were generated including distance to dunite, serpentine, and peridotite bodies as the evidence of host rock, the distance to the polygons of serpentinized alteration as the evidence of alteration, and the geochemical signature of Cr, Ni, and Co of stream sediment geochemical data as the geochemical evidence layers. The random forests algorithm was trained according to the values for evidence layers in the deposit and non-deposit locations as the positive and negative sites, respectively. The success- and prediction-rate curves were employed to assess the spatial correlation of the model with known deposits and to assess its predicting ability, respectively. The results obtained demonstrated that the model could be successfully applied to generate the exploration targets. The exploration targets occupied 19% of the studied area, and covered all the known podiform chromite deposits. Therefore, it is worthy to follow-up the detailed exploration surveys in the delineated target zones and follow the exploration licenses in delimited targets.

## Acknowledgments

The authors are appreciative of anonymous reviewers whose valuable comments and recommendations significantly improved the presentation of this work.

## References

- [1]. Bonham-Carter, G.F., Agterberg, F.P. and Wright, D.F. (1990). Weights of evidence modeling: a new approach to mapping mineral potential. Statistical applications in the earth sciences. 89: 171-183.
- [2]. Carranza, E.J.M. (2008). Geochemical anomaly and mineral prospectivity mapping in GIS. Vol. 11. Elsevier, Amsterdam.
- [3]. Ghavami-Riabi, R., Seyedrahimi-Niaraq, M.M., Khalokakaie, R. and Hazareh, M.R. (2010). U-spatial

statistic data modeled on a probability diagram for investigation of mineralization phases and exploration of shear zone gold deposits. Journal of Geochemical Exploration. 104: 27-33.

- [4]. Ziaii, M., Abedi, A. and Ziaei, M. (2009). Geochemical and mineralogical pattern recognition and modeling with a Bayesian approach to hydrothermal gold deposits. Applied Geochemistry. 24: 1142-1146.
- [5]. Ziaii, M., Pouyan, A.A. and Ziaei, M. (2009). Neuro-fuzzy modeling in mining geochemistry: identification of geochemical anomalies. Journal of Geochemical Exploration. 100: 25-36.
- [6]. Ziaii, M., Carranza, E.J.M. and Ziaei, M. (2011). Application of geochemical zonality coefficients in mineral prospectivity mapping. Computers & geosciences. 37:1935-1945.
- [7]. Ziaii, M., Ardejani, F.D., Ziaei, M. and Soleymani, A.A. (2012). Neuro-fuzzy modeling based genetic algorithms for identification of geochemical anomalies in mining geochemistry. Applied geochemistry. 27: 663-676.
- [8]. Navidi, A., Ziaii, M., Afzal, P., Yasrebi, A.B., Wetherelt, A. and Foster, P. (2014). Determination of Chromites Prospects Using Multifactorial Models and Zonality Index in the Parang 1: 100000 Sheet, Iran. Universal Journal of Geoscience. 2: 133-139.
- [9]. Momeni, S., Shahrokhi, S.V., Afzal, P., Sadeghi, B., Farhadinejad, T. and Nikzad, M.R. (2016). Delineation of the Cr mineralization based on the stream sediment data utilizing fractal modeling and factor analysis in the Khoy 1: 100,000 sheet, NW Iran. Bulletin of The Mineral Research and Exploration. 152: 143-151.
- [10]. Shahi, H., Ghavami, R. and Rouhani, A.K. (2016). Detection of deep and blind mineral deposits using new proposed frequency coefficients method in frequency domain of geochemical data. Journal of Geochemical Exploration. 162: 29-39.
- [11]. Afzal, P., Yasrebi, A.B., Saein, L.D. and Panahi, S. (2017). Prospecting of Ni mineralization based on geochemical exploration in Iran. Journal of Geochemical Exploration. 181: 294-304.
- [12]. Yousefi, M. and Carranza, E.J.M. (2015). Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. Computers & Geosciences. 74: 97-109.
- [13]. Yousefi, M. and Carranza, E.J.M. (2015). Prediction-area (P-A) plot and C-A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. Computers & Geosciences. 79: 69-81.
- [14]. Yousefi, M. and Carranza, E.J.M. (2016). Union score and fuzzy logic mineral prospectivity mapping using discretized and continuous spatial evidence values. Journal of African Earth Sciences. 128: 47-60.

- [15]. Parsa, M., Maghsoudi, A. and Yousefi, M. (2018). Spatial analyses of exploration evidence data to model skarn-type copper prospectivity in the Varzaghan district, NW Iran. *Ore Geology Reviews*. 92: 97-112.
- [16]. Parsa, M., Maghsoudi, A. and Yousefi, M. (2017). An improved data-driven fuzzy mineral prospectivity mapping procedure; cosine amplitude-based similarity approach to delineate exploration targets. *International Journal of Applied Earth Observation and Geoinformation*. 58: 157-167.
- [17]. Parsa, M. and Maghsoudi, A. (2018). Controls on Mississippi Valley-Type Zn-Pb mineralization in Behabad district, Central Iran: Constraints from spatial and numerical analyses. *Journal of African Earth Sciences*. 140: 189-198.
- [18]. Ford, A. and Hart, C.J. (2013). Mineral potential mapping in frontier regions: a Mongolian case study. *Ore Geology Reviews*. 51: 15-26.
- [19]. Carranza, E.J.M., Hale, M. and Faassen, C. (2008). Selection of coherent deposit-type locations and their application in data-driven mineral prospectivity mapping. *Ore Geology Reviews*. 33: 536-558.
- [20]. Agterberg, F.P. and Bonham-Carter, G.F. (2005). Measuring the performance of mineral-potential maps. *Natural Resources Research*. 14: 1-17.
- [21]. Nykänen, V., Lahti, I., Niiranen, T. and Korhonen, K. (2015). Receiver operating characteristics (ROC) as validation tool for prospectivity models-a magmatic Ni-Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geology Reviews*. 71: 853-860.
- [22]. Berman, M. (1977). Distance distributions associated with Poisson processes of geometric figures. *Journal of Applied Probability*. 14: 195-199.
- [23]. Cheng, Q., Xia, Q., Li, W., Zhang, S., Chen, Z., Zuo, R. and Wang, W. (2010). Density/area power-law models for separating multi-scale anomalies of ore and toxic elements in stream sediments in Gejiu mineral district, Yunnan Province, China. *Biogeosciences*. 7: 3019-3025.
- [24]. Lindsay, M., Aitken, A., Ford, A., Dentith, M., Hollis, J. and Tyler, I. (2016). Reducing subjectivity in multi-commodity mineral prospectivity analyses: modeling the west Kimberley, Australia. *Ore Geology Reviews*. 76: 395-413.
- [25]. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*. 71: 804-818.
- [26]. Breiman, L. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC, London.
- [27]. Breiman, L. (1996). Bagging predictors. *Machine learning*. 24:123-140.
- [28]. Breiman, L. (2001). Random forests. *Machine learning*. 45: 5-32.
- [29]. Carranza, E.J.M. and Laborte, A.G. (2015). Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random forest algorithm. *Ore Geology Reviews*. 71: 777-787.
- [30]. Carranza, E.J.M. and Laborte, A.G. (2015). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers & Geosciences*. 74: 60-70.
- [31]. Carranza, E.J.M. and Laborte, A.G. (2016). Data-driven predictive modeling of mineral prospectivity using random forest: a case study in Catanduanes Island (Philippines). *Natural Resources Research*. 25: 35-50.
- [32]. Azizan, H. and Naderi, N. (1998). 1:00,000 geological map of Dolatabad. Geological survey of Iran.
- [33]. Mohebi, A. (2003). Report of prospective zones of mineral prospectivity in southeastern Iran. Geological survey of Iran. (In Persian)
- [34]. Aghanabati, A. (2004). Geology of Iran. Geological survey of Iran, Tehran. (In Persian)
- [35]. Lipin, B.R. (1984). Chromite from the Blue Ridge Province of North Carolina. *American Journal of Science*. 284: 507-529.
- [36]. Wells, F.G., Carter, F.T. and Rynearson, G.A. (1965). Chromite deposits of Del Norte County, California. Division of Mines.
- [37]. Yigit, O. (2009). Mineral deposits of Turkey in relation to Tethyan metallogeny: implications for future mineral exploration. *Economic Geology*. 104: 19-51.
- [38]. Thayer, T.P. (1946). Preliminary chemical correlation of chromite with the containing rocks. *Economic Geology*. 41: 202-217.
- [39]. Wakabayashi, J., Ghatak, A. and Basu, A.R. (2010). Suprasubduction-zone ophiolite generation, emplacement, and initiation of subduction: A perspective from geochemistry, metamorphism, geochronology, and regional geology. *Bulletin Geological society of America* 122: 1548-1568.
- [40]. Allen, J.E. (1941). Geological investigation of the chromite deposits of California. California State printing Office.
- [41]. Maxwell, J.C. (1949). Some occurrences of chromite in New Caledonia. *Economic Geology*. 44: 525-544.
- [42]. Thompson, M. and Howarth, R.J. (1976). Duplicate analysis in geochemical practice. Part I.

Theoretical approach and estimation of analytical reproducibility. *Analyst*. 101: 690-698.

- [43]. Carranza, E.J.M. (2009). Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity. *Computers & Geosciences*. 35: 2032-2046.
- [44]. Filzmoser, P., Hron, K., Reimann, C. and Garrett, R. (2009). Robust factor analysis for compositional data. *Computers & Geosciences*. 35: 1854-1861.
- [45]. Parsa, M., Maghsoudi, A., Yousefi, M. and Sadeghi, M. (2016). Recognition of significant multi-element geochemical signatures of porphyry Cu deposits in Noghdouz area, NW Iran. *Journal of Geochemical Exploration*. 165: 111-124.
- [46]. Parsa, M., Maghsoudi, A., Yousefi, M. and Sadeghi, M. (2017). Multifractal analysis of stream sediment geochemical data: Implications for hydrothermal nickel prospecting in an arid terrain, eastern Iran. *Journal of Geochemical Exploration*. 181: 305-317.
- [47]. Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*. 35: 279-300.
- [48]. Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London. 416 P.
- [49]. Yousefi, M., Kamkar-Rouhani, A. and Carranza, E.J.M. (2014). Application of staged factor analysis and logistic function to create a fuzzy stream sediment geochemical evidence layer for mineral prospectivity mapping. *Geochemistry: Exploration, Environment, Analysis*. 14: 45-58.
- [50]. Liaw, A. and Wiener, M. (2002). Classification and regression by randomForest. *R news*. 2:18-22.
- [51]. R Development core team (2008). The R project for statistical computing. <http://www.R-project.org>.
- [52]. Chung, C.J.F. and Fabbri, A.G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*. 30: 451-472.
- [53]. Parsa, M., Maghsoudi, A., Yousefi, M. and Sadeghi, M. (2016). Prospectivity modeling of porphyry- Cu deposits by identification and integration of efficient mono-elemental geochemical signatures. *Journal of African Earth Sciences*. 114: 228-241.
- [54]. Mosier, D.L., Singer, D.A., Moring, B.C. and Galloway, J.P. (2012). Podiform chromite deposits-database and grade and tonnage models. United States Geological Survey.
- [55]. Kospiri, A. (1999). Case histories of the application of geophysical methods to chromite exploration in the balkans. Second Balkan Geophysical Congress and Exhibition, July 5-9, 1999, Istanbul: Balkan Geophysical Society.



## کاربرد الگوریتم جنگل های تصادفی در شناسایی مناطق مستعد کانی زایی کرومیت انبانه های در منطقه دولت آباد، جنوب شرقی ایران

محمدپارسا صدر\* و مصطفی ناظری

مرکز تحقیقات و خودکفایی قرارگاه سازندگی خاتم الانبیاء، تهران، ایران

ارسال ۲۰۱۸/۳/۴، پذیرش ۲۰۱۸/۳/۳۰

\* نویسنده مسئول مکاتبات: muhamadmilresearch@gmail.com

### چکیده:

منطقه دولت آباد در جنوب شرقی ایران واقع شده است و دارای تعدادی از کانی زایی های شناخته شده کرومیت است. این کانی زایی ها در زون آمیزه رنگین متشکل از هارزبوژیت، دونیت و پیروکسنیت دیده می شوند. کانی زایی های کرومیت در منطقه دولت آباد دارای اشکال نامنظم هستند و به صورت عدسی هایی در زون آمیزه رنگین توزیع شده اند. منطقه دولت آباد واجد پتانسیل بسیار خوبی به منظور کشف و شناسایی کانی زایی های بیشتری از کرومیت است؛ بنابراین، هدف از این مطالعه، شناسایی مناطق مستعد کانی زایی کرومیت انبانه های برای تمرکز برنامه های اکتشافی است. به منظور دستیابی به هدف مذکور، الگوریتم جنگل های تصادفی برای تلفیق اثرهای مرتبط با کانی زایی کرومیت در داده های اکتشافی به کار گرفته شد. ویژگی های ژنتیکی کانسارهای کرومیت انبانه های برای تعریف معیارهای اکتشافی استفاده شدند. سپس این معیارها، به لایه های شاهد اکتشافی تبدیل شدند. لایه های شاهد اکتشافی مؤثر، یعنی لایه های دارای ارتباط مثبت با کانی زایی، با استفاده از تحلیل مسافت- فاصله شناسایی و از لایه های شاهد غیر مؤثر تفکیک شدند. با توجه به موقعیت کانی زایی های کرومیت انبانه های شناخته شده در منطقه و لایه های شاهد اکتشافی مؤثر، یک مدل پیشگوی جنگل های تصادفی آموزش دیده شد و به منظور شناسایی مناطق مستعد کانی زایی کرومیت انبانه های استفاده شد. اهداف اکتشافی شناسایی شده ۱۹٪ از منطقه دولت آباد را که شامل ۱۰۰٪ کانی زایی های شناخته شده کرومیت انبانه های در منطقه می باشند، تحت پوشش قرار می دهند؛ بنابراین، انجام فعالیت های اکتشافی تفصیلی در مناطق تعیین مرز شده سودمند است.

**کلمات کلیدی:** کانسارهای کرومیت انبانه های، جنگل های تصادفی، پتانسیل یابی مواد معدنی.